Haplotype mapping in pharmacogenetics

1. The identification of tagging SNPs in ADME genes
2. An application of haplotype tagging

What is a tagging SNP anyway?

In this (unrealistic!) case the five haplotypes can be represented by just five tagging SNPs

---

A paradigm for indirect association studies (HapMap)

- Genotype "control" individuals (LD data set)
- Assess LD and select tSNPs
- Type tSNPs in "phenotyped" individuals

Questions about ht mapping

UCL

- **How to select and test performance of tSNPs?**
- **How many samples are required to select tSNPs?**
- **How does performance depend on MAF?**
- **How does performance depend on density?**
- **Is functional variation represented as well as non functional variation?**
- **How many more tSNPs are required to perform well in multiple populations?**

Selection of Tags

UCL

- *Select "tagging" SNPs that define G haplotypes that satisfy performance criteria in a regression model in which the G haplotypes are used as independent predictors to predict the allelic states of other "tagged" SNPs.*

- $Y_i = x_{i1}b_1 + x_{i2}b_2 + \ldots + x_{iG}b_G$

Refs: Weale ME et al. (2003) *American Journal of Human Genetics* **73**: 551-565

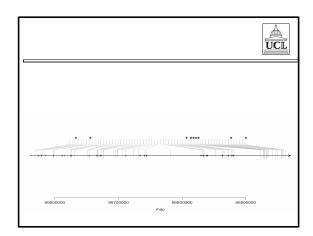Goldstein DB et al. (2003) *Trends in Genetics* **19**: 615-622

Software implementation:

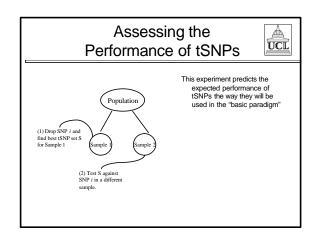Weale ME & Goldstein DB (2003) *TagIT software*, http://popgen.biol.ucl.ac.uk
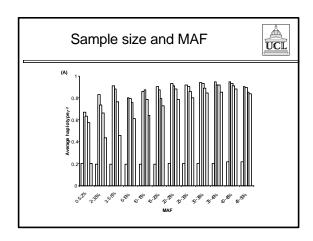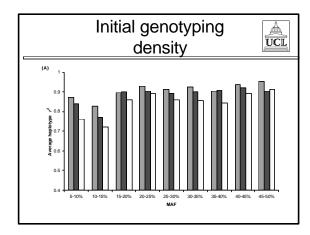
The data set …

UCL

- 60 genes encoding most of the important DMEs
- Genotype approximately 1,000 SNPs in 64 CEPH and 64 Japanese individuals for a target density of one SNP / 2kb

    -- GSK / UCL Collaboration --

Assessing the Performance of tSNPs

This experiment predicts the expected performance of tSNPs the way they will be used in the "basic paradigm"

Population

(1) Drop SNP i and find best tSNP set S for Sample 1

Sample 1   Sample 2

(2) Test S against SNP i in a different sample.

Sample size and MAF

UCL

(A)

Average haplotype r²

0.5-2% 2-3% 3.5-5% 5-10% 10-15% 15-20% 20-25% 25-30% 30-35% 35-40% 40-45% 45-50%

MAF

Initial genotyping density

UCL

(A)

Average haplotype r²

5-10% 10-15% 15-20% 20-25% 25-30% 30-35% 35-40% 40-45% 45-50%

MAF

Cosmopolitan SNPs

UCL

- Europe   196
- Japan   179

Cosmopolitan SNPs

- Europe   196
- Japan   179

- Europe and Japan   226

Representing functional variation

Will tags hold in a different population

- Tags defined in CEPH
- Genotyped in a large sample from Aberdeen
- And tested against the panel of functional variants

Application in a new population

Tagging the major human DMEs

- **Density > 1 SNP / 5 kb results in detectable (but modest?) improvement in power**
- **Cosmopolitan tSNP sets require only modest increases in numbers over single population sets**
- **Functional SNPs can be represented as well as non functional ones**

Tagging the major human DMEs

- **Density > 1 SNP / 5 kb results in detectable (but modest?) improvement in power**
- **Cosmopolitan tSNP sets require only modest increases in numbers over single population sets**
- **Functional SNPs can be represented as well as non functional ones**
- **SNPs with low MAF are trouble!**

Tagging the major human DMEs

UCL

- **Approximately 200 SNPs are sufficient to represent the greater than 4,000 common polymorphisms in key genes regulating drug plasma levels**

Phenytoin

UCL
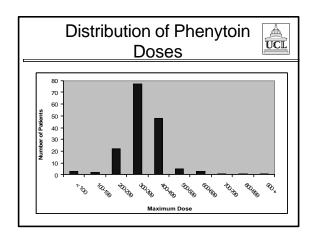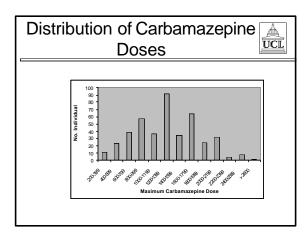
- A once widely used and effective anti epileptic drug that is now less often used (especially in Europe), in part because of adverse reactions (gum hyptertrophy, vision, headache)

Carbamazepine

UCL
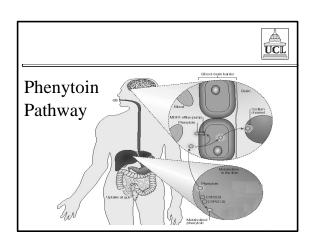
- Overall a safer drug than PHT, and more commonly used in the UK
- Identification of therapeutic dose can take months, and is complicated by tolerance

Distribution of Phenytoin Doses

Distribution of Carbamazepine Doses

Phenytoin Pathway
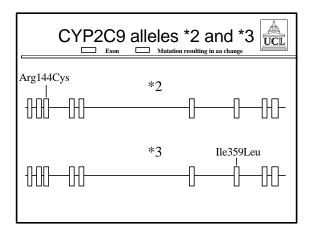
Strategy
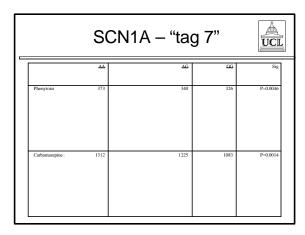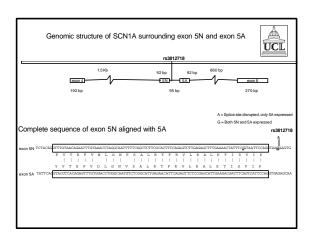
Phenytoin
- Assess functional variants in CYP2C9
- Assess putative functional variant in ABCB1
- Assess tagging SNPs in SCN1A

Carbamazepine
- Assess putative functional variant in ABCB1
- Assess tagging SNPs in SCN1A
- *Assess variation in PXR, regulator of CYP3A4*

CYP2C9 alleles *2 and *3

⬜ **Exon**   ⬜ **Mutation resulting in aa change**

Arg144Cys

*2

*3   Ile359Leu

Representing variatiion in SCN1A

Gene	Size (Kb)	# SNPs	Density	# Tags
SCN1a	139	15	9.26667	4

Results

CYP2C9 Allele	Freq. (%)	Average Phenytoin Dose (mg)			Sig.
		wt/wt	wt/var	var/var	
*2	11.9	357	339	317	0.38
*3	9.7	362	304	250	**0.005**

SCN1A – "tag 7"

	AA	AG	GG	Sig
Phenytoin	373	340	326	P=0.0046
Carbamazepine :	1312	1225	1083	P=0.0014

Genomic structure of SCN1A surrounding exon 5N and exon 5A

rs3812718

A = Splice site disrupted, only 5A expressed
G = Both 5N and 5A expressed

Complete sequence of exon 5N aligned with 5A

Some good news (finally)…

UCL

- Haplotype tagging can identify unknown functional variants

PHT and CBZ dosing

UCL

- PHT: Combining SCN1A and CYP2C9, the average maximum dose range between genetically sub-defined groups is 130 mg
- For CBZ, the average maximum dose range across SC1NA genotypes is 300 mg per day.

Clinical relevance?

UCL

- Rate at which AED dose can be safely increased
- Only prospective trials can determine whether PGx can improve clinical use will determine whether the clinical use of these medicines

Pgx is a simpler complex trait

UCL

- Obvious candidate genes often carry PGx variants (drug specific)
- There may be direct clinical importance, although the demonstration will generally require prospective trials

Acknowledgments

UCL

- Tagging human drug metabolism
  – GlaxoSmithKline and University College London
- Epilepsy genetics
  – Gianpierro Cavalleri, Chantal Depondt, Stephanie Schorge, Nicole Soranzo, Asra Siddiqui, Mike Weale,, Chantal Depondt, Ley Sander, Mari Wyn Burley, Richard Marguerie, Nicholas Wood, Sanjay Sisodiya